

# From $f$ -divergence to quantum quasi-entropies and their use

*Dedicated to Professor Imre Csiszár*

Dénes Petz<sup>1</sup>

Alfréd Rényi Institute of Mathematics, H-1364 Budapest, POB 127, Hungary

## Abstract

Csiszár's  $f$ -divergence of two probability distributions was extended to the quantum case by the author in 1985. In the quantum setting positive semidefinite matrices are in the place of probability distributions and the quantum generalization is called quasi-entropy which is related to some other important concepts as covariance, quadratic costs, Fisher information, Cramér-Rao inequality and uncertainty relation. A conjecture about the scalar curvature of a Fisher information geometry is explained. The described subjects are overviewed in details in the matrix setting, but at the very end the von Neumann algebra approach is sketched shortly.

*Key words and phrases:*  $f$ -divergence, quasi-entropy, von Neumann entropy, relative entropy, monotonicity property, Fisher information, uncertainty.

Let  $\mathcal{X}$  be a finite space with probability measures  $p$  and  $q$ . Their **relative entropy** or **divergence**

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

was introduced by Kullback and Leibler in 1951 [27]. More precisely, if  $p(x) = q(x) = 0$ , then  $\log(p(x)/q(x)) = 0$  and if  $p(x) \neq 0$  but  $q(x) = 0$  for some  $x \in \mathcal{X}$ , then  $\log(p(x)/q(x)) = +\infty$ .

A possible generalization of the relative entropy is the  $f$ -divergence introduced by Csiszár:

$$D_f(p||q) = \sum_{x \in \mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right) \quad (1)$$

---

<sup>1</sup>E-mail: petz@math.bme.hu. Partially supported by the Hungarian Research Grant OTKA T068258.

with a real function  $f(x)$  defined for  $x > 0$  [6, 8]. For the convex function  $f(x) = x \log x$  the relative entropy is obtained.

This paper first gives a rather short survey about  $f$ -divergence and we turn to the non-commutative (algebraic, or quantum) generalization. Roughly speaking this means that the positive  $n$ -tuples  $p$  and  $q$  are replaced by positive semidefinite  $n \times n$  matrices and the main questions in the study remain rather similar to the probabilistic case. The quantum generalization was called quasi-entropy and it is related to some other important concepts as covariance, quadratic costs, Fisher information, Cramér-Rao inequality and uncertainty relation. These subjects are overviewed in details in the matrix setting, but at the very end the von Neumann algebra approach is sketched shortly. When the details are not presented, the precise references are given.

## 1 $f$ -divergence and its use

Let  $\mathcal{F}$  be the set of continuous convex functions  $\mathbb{R}^+ \rightarrow \mathbb{R}$ . The following result explains the importance of convexity.

Let  $\mathcal{A}$  be a partition of  $\mathcal{X}$ . If  $p$  is a probability distribution on  $\mathcal{X}$ , then  $p_{\mathcal{A}}(A) := \sum_{x \in A} p(x)$  becomes a probability distribution on  $\mathcal{A}$

**Theorem 1** *Let  $\mathcal{A}$  be a partition of  $\mathcal{X}$  and  $p, q$  be probability distributions on  $\mathcal{X}$ . If  $f \in \mathcal{F}$ , then*

$$D_f(p_{\mathcal{A}} || q_{\mathcal{A}}) \leq D_f(p || q).$$

The inequality in the theorem is the monotonicity of the  $f$ -divergence. A particular case is

$$f(1) \leq D_f(p || q).$$

**Theorem 2** *Let  $f, g \in \mathcal{F}$  and assume that*

$$D_f(p || q) = D_g(p || q).$$

*for every distribution  $p$  and  $q$ . Then there exists a constant  $c \in \mathbb{R}$  such that  $f(x) - g(x) = c(x - 1)$ .*

Since the divergence is a kind of informational distance, we want  $D_f(p || p) = 0$  and require  $f(1) = 0$ . This is nothing else but a normalization,

$$D_{f+c}(p || q) = D_f(p || q) + c.$$

A bit more generally, we can say that if  $f(x) - g(x)$  is a linear function, then  $D_f$  and  $D_g$  are essentially the same quantities.

It is interesting to remark that  $qf(p/q)$  can be considered also as a mean of  $p$  and  $q$ . In that case the mean of  $p$  and  $p$  should be  $p$ , so in the theory of means  $f(1) = 1$  is a different natural requirement.

Set  $f^*(x) = xf(x^{-1})$ . Then  $D_f(p||q) = D_{f^*}(q||p)$ . The equality  $f^* = f$  is the symmetry condition.

**Example 1** Let  $f(x) = |x - 1|$ . Then

$$D_f(p, q) = \sum_x |p(x) - q(x)| =: V(p, q)$$

is the **variational distance** of  $p$  and  $q$ .  $\square$

**Example 2** Let  $f(x) = (1 - \sqrt{x})^2/2$ . Then

$$D_f(p, q) = \sum_x (\sqrt{p(x)} - \sqrt{q(x)})^2 =: H^2(p, q)$$

is the **squared Hellinger distance** of  $p$  and  $q$ .  $\square$

**Example 3** The function

$$f_\alpha(t) = \frac{1}{\alpha(1-\alpha)}(1-t^\alpha)$$

gives the relative  $\alpha$ -entropy

$$S_\alpha(p||q) = \frac{1}{\alpha(1-\alpha)} \left( 1 - \sum_x p(x)^\alpha q(x)^{1-\alpha} \right). \quad (2)$$

The limit  $\alpha \rightarrow 0$  gives the relative entropy.  $\square$

Several other functions appeared in the literature, we list a few of them:

$$f^{(s)}(x) = \frac{1}{s(1-s)}(1+x-x^s-x^{1-s}) \quad 0 < s \neq 1 \quad [9], \quad (3)$$

$$f_\beta(x) = \begin{cases} \frac{1}{1-1/\beta} ((1+x^\beta)^{1/\beta} - 2^{1/\beta-1}(1+x)) & \text{if } 0 < \beta \neq 1, \\ (1+x)\log 2 + x\log x - (1+x)\log(x+1) & \text{if } \beta = 1. \end{cases} \quad [32] \quad (4)$$

The following result of Csiszár is a characterization (or axiomatization) of the  $f$ -divergence.

**Theorem 3** Assume that a number  $C(p, q) \in \mathbb{R}$  is associated to probability distributions on the same set  $\mathcal{X}$  for all finite sets  $\mathcal{X}$ . If

- (a)  $C(p, q)$  is invariant under the permutations of the basic set  $\mathcal{X}$ .

(b) if  $\mathcal{A}$  is a partition of  $\mathcal{X}$ , then  $C(p_{\mathcal{A}}, q_{\mathcal{A}}) \leq C(p, q)$  and the equality holds if and only if

$$p_{\mathcal{A}}(A)q(x) = q_{\mathcal{A}}(A)p(x)$$

whenever  $x \in A \in \mathcal{A}$ ,

then there exists a convex function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  which is continuous at 0 and  $C(p, q) = D_f(p||q)$  for every  $p$  and  $q$ .

## 2 Quantum quasi-entropy

In the mathematical formalism of quantum mechanics, instead of  $n$ -tuples of numbers one works with  $n \times n$  complex matrices. They form an algebra and this allows an algebraic approach. In this approach, a probability density is replaced by a positive semidefinite matrix of trace 1 which is called **density matrix**[39]. The eigenvalues of a density matrix give a probability density. However, this is not the only probability density provided by a density matrix. If we rewrite the matrix in a certain orthonormal basis, then the diagonal element  $p_1, p_2, \dots, p_n$  form a probability density.

Let  $\mathcal{M}$  denote the algebra of  $n \times n$  matrices with complex entries. For positive definite matrices  $\rho_1, \rho_2 \in \mathcal{M}$ , for  $A \in \mathcal{M}$  and a function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ , the **quasi-entropy** is defined as

$$\begin{aligned} S_f^A(\rho_1 \parallel \rho_2) &:= \langle A\rho_2^{1/2}, f(\Delta(\rho_1/\rho_2))(A\rho_2^{1/2}) \rangle \\ &= \text{Tr } \rho_2^{1/2} A^* f(\Delta(\rho_1/\rho_2))(A\rho_2^{1/2}), \end{aligned} \quad (5)$$

where  $\langle B, C \rangle := \text{Tr } B^* C$  is the so-called **Hilbert-Schmidt inner product** and  $\Delta(\rho_1/\rho_2) : \mathcal{M} \rightarrow \mathcal{M}$  is a linear mapping acting on matrices:

$$\Delta(\rho_1/\rho_2)A = \rho_1 A \rho_2^{-1}.$$

This concept was introduced in [33, 34], see also Chapter 7 in [31] and it is the quantum generalization of the  $f$ -entropy of Csiszár used in classical information theory (and statistics) [7, 30].

The monotonicity in Theorem 1 is the consequence of the Jensen inequality. A function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is called **matrix concave** if one of the following two equivalent conditions holds:

$$f(\lambda A + (1 - \lambda)B) \geq \lambda f(A) + (1 - \lambda)f(B) \quad (6)$$

for every number  $0 < \lambda < 1$  and for positive definite square matrices  $A$  and  $B$  (of the same size). In the other condition the number  $\lambda$  is (heuristically) replaced by a matrix:

$$f(CAC^* + DBD^*) \geq Cf(A)C^* + Df(B)D^* \quad (7)$$

if  $CC^* + DD^* = I$ .

A function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is called **matrix monotone** if for positive definite matrices  $A \leq B$  the inequality  $f(A) \leq f(B)$  holds. It is interesting that a matrix monotone function is matrix concave and a matrix concave function is matrix monotone if it is bounded from below [17].

Let  $\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$  be a mapping between two matrix algebras. The dual  $\alpha^* : \mathcal{M} \rightarrow \mathcal{M}_0$  with respect to the Hilbert-Schmidt inner product is positive if and only if  $\alpha$  is positive. Moreover,  $\alpha$  is unital if and only if  $\alpha^*$  is trace preserving.  $\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$  is called a **Schwarz mapping** if

$$\alpha(B^*B) \geq \alpha(B^*)\alpha(B) \quad (8)$$

for every  $B \in \mathcal{M}_0$ .

The quasi-entropies are monotone and jointly convex [31, 34].

**Theorem 4** *Assume that  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is an operator monotone function with  $f(0) \geq 0$  and  $\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$  is a unital Schwarz mapping. Then*

$$S_f^A(\alpha^*(\rho_1), \alpha^*(\rho_2)) \geq S_f^{\alpha(A)}(\rho_1, \rho_2) \quad (9)$$

holds for  $A \in \mathcal{M}_0$  and for invertible density matrices  $\rho_1$  and  $\rho_2$  from the matrix algebra  $\mathcal{M}$ .

*Proof:* The proof is based on inequalities for operator monotone and operator concave functions. First note that

$$S_{f+c}^A(\alpha^*(\rho_1), \alpha^*(\rho_2)) = S_f^A(\alpha^*(\rho_1), \alpha^*(\rho_2)) + c \operatorname{Tr} \rho_1 \alpha(A^* A)$$

and

$$S_{f+c}^{\alpha(A)}(\rho_1, \rho_2) = S_f^{\alpha(A)}(\rho_1, \rho_2) + c \operatorname{Tr} \rho_1 (\alpha(A)^* \alpha(A))$$

for a positive constant  $c$ . Due to the Schwarz inequality (8), we may assume that  $f(0) = 0$ .

Let  $\Delta := \Delta(\rho_1/\rho_2)$  and  $\Delta_0 := \Delta(\alpha^*(\rho_1)/\alpha^*(\rho_2))$ . The operator

$$V X \alpha^*(\rho_2)^{1/2} = \alpha(X) \rho_2^{1/2} \quad (X \in \mathcal{M}_0) \quad (10)$$

is a contraction:

$$\begin{aligned} \|\alpha(X) \rho_2^{1/2}\|^2 &= \operatorname{Tr} \rho_2 (\alpha(X)^* \alpha(X)) \\ &\leq \operatorname{Tr} \rho_2 (\alpha(X^* X)) = \operatorname{Tr} \alpha^*(\rho_2) X^* X = \|X \alpha^*(\rho_2)^{1/2}\|^2 \end{aligned}$$

since the Schwarz inequality is applicable to  $\alpha$ . A similar simple computation gives that

$$V^* \Delta V \leq \Delta_0. \quad (11)$$

Since  $f$  is operator monotone, we have  $f(\Delta_0) \geq f(V^*\Delta V)$ . Recall that  $f$  is operator concave, therefore  $f(V^*\Delta V) \geq V^*f(\Delta)V$  and we conclude

$$f(\Delta_0) \geq V^*f(\Delta)V. \quad (12)$$

Application to the vector  $A\alpha^*(\rho_2)^{1/2}$  gives the statement.  $\square$

It is remarkable that for a multiplicative  $\alpha$  we do not need the condition  $f(0) \geq 0$ . Moreover,  $V^*\Delta V = \Delta_0$  and we do not need the matrix monotonicity of the function  $f$ . In this case the only condition is the matrix concavity, analogously to Theorem 1.

If we apply the monotonicity (9) to the embedding  $\alpha(X) = X \oplus X$  of  $\mathcal{M}$  into  $\mathcal{M} \oplus \mathcal{M}$  and to the densities  $\rho_1 = \lambda E_1 \oplus (1 - \lambda)F_1$ ,  $\rho_2 = \lambda E_2 \oplus (1 - \lambda)F_2$ , then we obtain the joint concavity of the quasi-entropy:

$$\lambda S_f^A(E_1, E_2) + (1 - \lambda)S_f^A(F_1, F_2) \leq S_f^A(\lambda E_1 + (1 - \lambda)E_2) + S_f^A(\lambda F_1 + (1 - \lambda)F_2)$$

holds. The case  $f(t) = t^\alpha$  is the famous Lieb's concavity theorem:  $\text{Tr } A\rho^\alpha A^* \rho^{1-\alpha}$  is concave in  $\rho$  [29].

The concept of quasi-entropy includes some important special cases. If  $\rho_2$  and  $\rho_1$  are different and  $A = I$ , then we have a kind of relative entropy. For  $f(x) = x \log x$  we have Umegaki's relative entropy  $S(\rho_1 \| \rho_2) = \text{Tr } \rho_1 (\log \rho_1 - \log \rho_2)$ . (If we want a matrix monotone function, then we can take  $f(x) = \log x$  and then we get  $S(\rho_2 \| \rho_1)$ .) Umegaki's relative entropy is the most important example, therefore the function  $f$  will be chosen to be matrix convex. This makes the probabilistic and non-commutative situation compatible as one can see in the next argument.

Let  $\rho_1$  and  $\rho_2$  be density matrices in  $\mathcal{M}$ . If in certain basis they have diagonal  $p = (p_1, p_2, \dots, p_n)$  and  $q = (q_1, q_2, \dots, q_n)$ , then the monotonicity theorem gives the inequality

$$D_f(p \| q) \leq S_f(\rho_1 \| \rho_2) \quad (13)$$

for a matrix convex function  $f$ . If  $\rho_1$  and  $\rho_2$  commute, them we can take the common eigenbasis and in (13) the equality appears. It is not trivial that otherwise the inequality is strict.

If  $\rho_1$  and  $\rho_2$  are different, then there is a choice for  $p$  and  $q$  such that they are different as well. Then

$$0 < D_f(p \| q) \leq S_f(\rho_1 \| \rho_2).$$

Conversely, if  $S_f(\rho_1 \| \rho_2) = 0$ , then  $p = q$  for every basis and this implies  $\rho_1 = \rho_2$ . For the relative entropy, a deeper result is known. The **Pinsker-Csiszár inequality** says that

$$(\|p - q\|_1)^2 \leq 2D(p \| q). \quad (14)$$

This extends to the quantum case as

$$(\|\rho_1 - \rho_2\|_1)^2 \leq 2S(\rho_1 \| \rho_2), \quad (15)$$

see [22], or [39, Chap. 3].

**Problem 1** It would be interesting to extend Theorem 3 of Csiszár to the quantum case. If we require monotonicity and specify the condition for equality, then a function  $f$  is provided by Theorem 3, but for non-commuting densities the conclusion is not clear.

**Example 4** Let

$$f_\alpha(x) = \frac{1}{\alpha(1-\alpha)}(1-x^\alpha),$$

is matrix monotone decreasing for  $\alpha \in (-1, 1)$ . (For  $\alpha = 0$ , the limit is taken and it is  $-\log x$ .) Then the **relative entropies of degree  $\alpha$**  are produced:

$$S_\alpha(\rho_2 \parallel \rho_1) := \frac{1}{\alpha(1-\alpha)} \text{Tr} (I - \rho_1^\alpha \rho_2^{-\alpha}) \rho_2.$$

These quantities are essential in the quantum case.  $\square$

If  $\rho_2 = \rho_1 = \rho$  and  $A, B \in \mathcal{M}$  are arbitrary, then one can approach to the **generalized covariance** [38].

$$\text{qCov}_\rho^f(A, B) := \langle A\rho^{1/2}, f(\Delta(\rho/\rho))(B\rho^{1/2}) \rangle - (\text{Tr } \rho A^*)(\text{Tr } \rho B). \quad (16)$$

is a generalized covariance. If  $\rho, A$  and  $B$  commute, then this becomes  $f(1)\text{Tr } \rho A^* B - (\text{Tr } \rho A^*)(\text{Tr } \rho B)$ . This shows that the normalization  $f(1) = 1$  is natural. The generalized covariance  $\text{qCov}_\rho^f(A, B)$  is a sesquilinear form and it is determined by  $\text{qCov}_\rho^f(A, A)$  when  $\{A \in \mathcal{M} : \text{Tr } \rho A = 0\}$ . Formally, this is a quasi-entropy and Theorem 4 applies if  $f$  is matrix monotone. If we require the symmetry condition  $\text{qCov}_\rho^f(A, A) = \text{qCov}_\rho^f(A^*, A^*)$ , then  $f$  should have the symmetry  $xf(x^{-1}) = f(x)$ .

Assume that  $\text{Tr } \rho A = \text{Tr } \rho B = 0$  and  $\rho = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ . Then

$$\text{qCov}_\rho^f(A, B) = \sum_{ij} \lambda_i f(\lambda_j/\lambda_i) A_{ij}^* B_{ij}. \quad (17)$$

A matrix monotone function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  will be called **standard** if  $xf(x^{-1}) = f(x)$  and  $f(1) = 1$ . A standard function  $f$  admits a canonical representation

$$f(t) = \frac{1+t}{2} \exp \int_0^1 (1-t^2) \frac{\lambda-1}{(\lambda+t)(1+\lambda t)(\lambda+1)} h(\lambda) d\lambda, \quad (18)$$

where  $h : [0, 1] \rightarrow [0, 1]$  is a measurable function [18].

The usual **symmetrized covariance** corresponds to the function  $f(t) = (t+1)/2$ :

$$\text{Cov}_\rho(A, B) := \frac{1}{2} \text{Tr} (\rho(A^* B + B A^*)) - (\text{Tr } \rho A^*)(\text{Tr } \rho B).$$

The interpretation of the covariances is not at all clear. In the next section they will be called **quadratic cost functions**. It turns out that there is a one-to-one correspondence between quadratic cost functions and Fisher informations.

### 3 Fisher information

#### 3.1 The Cramér-Rao inequality

The Cramér-Rao inequality belongs to the basics of estimation theory in mathematical statistics. Its quantum analog was discovered immediately after the foundation of mathematical quantum estimation theory in the 1960's, see the book [21] of Helstrom, or the book [24] of Holevo for a rigorous summary of the subject. Although both the classical Cramér-Rao inequality and its quantum analog are as trivial as the Schwarz inequality, the subject takes a lot of attention because it is located on the highly exciting boundary of statistics, information and quantum theory.

As a starting point we give a very general form of the quantum Cramér-Rao inequality in the simple setting of finite dimensional quantum mechanics. For  $\theta \in (-\varepsilon, \varepsilon) \subset \mathbb{R}$  a statistical operator  $\rho(\theta)$  is given and the aim is to estimate the value of the parameter  $\theta$  close to 0. Formally  $\rho(\theta)$  is an  $n \times n$  positive semidefinite matrix of trace 1 which describes a mixed state of a quantum mechanical system and we assume that  $\rho(\theta)$  is smooth (in  $\theta$ ). Assume that an estimation is performed by the measurement of a self-adjoint matrix  $A$  playing the role of an observable.  $A$  is called **locally unbiased estimator** if

$$\frac{\partial}{\partial \theta} \text{Tr } \rho(\theta) A \Big|_{\theta=0} = 1. \quad (19)$$

This condition holds if  $A$  is an unbiased estimator for  $\theta$ , that is

$$\text{Tr } \rho(\theta) A = \theta \quad (\theta \in (-\varepsilon, \varepsilon)). \quad (20)$$

To require this equality for all values of the parameter is a serious restriction on the observable  $A$  and we prefer to use the weaker condition (19).

Let  $\varphi_0[K, L]$  be an inner product (or quadratic cost function) on the linear space of self-adjoint matrices. When  $\rho(\theta)$  is smooth in  $\theta$ , as already was assumed above, then

$$\frac{\partial}{\partial \theta} \text{Tr } \rho(\theta) B \Big|_{\theta=0} = \varphi_0[B, L] \quad (21)$$

with some  $L = L^*$ . From (19) and (21), we have  $\varphi_0[A, L] = 1$  and the Schwarz inequality yields

$$\varphi_0[A, A] \geq \frac{1}{\varphi_0[L, L]}. \quad (22)$$

This is the celebrated **inequality of Cramér-Rao type** for the locally unbiased estimator.

The right-hand-side of (22) is independent of the estimator and provides a lower bound for the quadratic cost. The denominator  $\varphi_0[L, L]$  appears to be in the role of Fisher information here. We call it quantum Fisher information with respect to the

cost function  $\varphi_0[\cdot, \cdot]$ . This quantity depends on the tangent of the curve  $\rho(\theta)$ . If the densities  $\rho(\theta)$  and the estimator  $A$  commute, then

$$L = \rho_0^{-1} \frac{d\rho(\theta)}{d\theta} \quad \text{and} \quad \varphi_0[L, L] = \text{Tr } \rho_0^{-1} \left( \frac{d\rho(\theta)}{d\theta} \right)^2 = \text{Tr } \rho_0 \left( \rho_0^{-1} \frac{d\rho(\theta)}{d\theta} \right)^2. \quad (23)$$

We want to conclude from the above argument that whatever Fisher information and generalized variance are in the quantum mechanical setting, they are very strongly related. In an earlier work [36, 37] we used a monotonicity condition to make a limitation on the class of Riemannian metrics on the state space of a quantum system. The monotone metrics are called Fisher information quantities in this paper.

Since the sufficient and necessary condition for the equality in the Schwarz inequality is well-known, we are able to analyze the case of equality in (22). The condition for equality is

$$A = \lambda L$$

for some constant  $\lambda \in \mathbb{R}$ . Therefore the necessary and sufficient condition for equality in (22) is

$$\dot{\rho}_0 := \frac{\partial}{\partial \theta} \rho(\theta) \Big|_{\theta=0} = \lambda^{-1} \mathbb{J}_0(A). \quad (24)$$

Therefore there exists a unique locally unbiased estimator  $A = \lambda \mathbb{J}_0^{-1}(\dot{\rho}_0)$ , where the number  $\lambda$  is chosen such a way that the condition (19) should be satisfied.

**Example 5** Let

$$\rho(\theta) := \rho + \theta B,$$

where  $\rho$  is a positive definite density and  $B$  is a self-adjoint traceless operator.  $A$  is locally unbiased when  $\text{Tr } AB = 1$ . In particular,

$$A = \frac{B}{\text{Tr } B^2}$$

is a locally unbiased estimator and in the Cramér-Rao inequality (22) the equality holds when  $\varphi_0[X, Y] = \text{Tr } XY$ , that is,  $\mathbb{J}_0$  is the identity.

If  $\text{Tr } \rho B = 0$  holds in addition, then the estimator is unbiased.  $\square$

## 3.2 Coarse-graining and monotonicity

In the simple setting in which the state is described by a density matrix, a **coarse-graining** is an affine mapping sending density matrices into density matrices. Such a mapping extends to all matrices and provides a positivity and trace preserving linear transformation. A common example of coarse-graining sends the density matrix  $\rho_{12}$  of a composite system  $1 + 2$  into the (reduced) density matrix  $\rho_1$  of component 1. There are several reasons to assume completely positivity about a coarse graining and we do so.

Assume that  $\rho(\theta)$  is a smooth curve of density matrices with tangent  $A := \dot{\rho}$  at  $\rho$ . The quantum Fisher information  $F_\rho(A)$  is an information quantity associated with the pair  $(\rho, A)$ , it appeared in the Cramér-Rao inequality above and the classical Fisher information gives a bound for the variance of a locally unbiased estimator. Let now  $\beta$  be a coarse-graining. Then  $\beta(\rho(\theta))$  is another curve in the state space. Due to the linearity of  $\beta$ , the tangent at  $\beta(\rho_0)$  is  $\beta(A)$ . As it is usual in statistics, information cannot be gained by coarse graining, therefore we expect that the Fisher information at the density matrix  $\rho_0$  in the direction  $A$  must be larger than the Fisher information at  $\beta(\rho_0)$  in the direction  $\beta(A)$ . This is the monotonicity property of the Fisher information under coarse-graining:

$$F_\rho(A) \geq F_{\beta(\rho)}(\beta(A)) \quad (25)$$

Although we do not want to have a concrete formula for the quantum Fisher information, we require that this monotonicity condition must hold. Another requirement is that  $F_\rho(A)$  should be quadratic in  $A$ , in other words there exists a non-degenerate real bilinear form  $\gamma_\rho(A, B)$  on the self-adjoint matrices such that

$$F_\rho(A) = \gamma_\rho(A, A). \quad (26)$$

The requirements (25) and (26) are strong enough to obtain a reasonable but still wide class of possible quantum Fisher informations.

We may assume that

$$\gamma_\rho(A, B) = \text{Tr } A \mathbb{J}_\rho^{-1}(B^*). \quad (27)$$

for an operator  $\mathbb{J}_\rho$  acting on matrices. (This formula expresses the inner product  $\gamma_D$  by means of the Hilbert-Schmidt inner product and the positive linear operator  $\mathbb{J}_\rho$ .) In terms of the operator  $\mathbb{J}_\rho$  the monotonicity condition reads as

$$\beta^* \mathbb{J}_{\beta(\rho)}^{-1} \beta \leq \mathbb{J}_\rho^{-1} \quad (28)$$

for every coarse graining  $\beta$ . ( $\beta^*$  stand for the adjoint of  $\beta$  with respect to the Hilbert-Schmidt product. Recall that  $\beta$  is completely positive and trace preserving if and only if  $\beta^*$  is completely positive and unital.) On the other hand the latter condition is equivalent to

$$\beta \mathbb{J}_\rho \beta^* \leq \mathbb{J}_{\beta(\rho)}. \quad (29)$$

We proved the following theorem in [36].

**Theorem 5** *If for every invertible density matrix  $\rho \in M_n(\mathbb{C})$  a positive definite sesquilinear form  $\gamma_\rho : M_n(\mathbb{C}) \times M_n(\mathbb{C}) \rightarrow \mathbb{C}$  is given such that*

*(1) the monotonicity*

$$\gamma_\rho(A, A) \geq \gamma_{\beta(\rho)}(\beta(A), \beta(A))$$

*holds for all completely positive coarse grainings  $\beta : M_n(\mathbb{C}) \rightarrow M_m(\mathbb{C})$ ,*

*(2)  $\gamma_\rho(A, A)$  is continuous in  $\rho$  for every fixed  $A$ ,*

$$(3) \quad \gamma_\rho(A, A) = \gamma_\rho(A^*, A^*),$$

$$(4) \quad \gamma_\rho(A, A) = \text{Tr } \rho^{-1} A^2 \text{ if } A \text{ is self-adjoint and } A\rho = \rho A,$$

then there exists a unique standard operator monotone function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  such that

$$\gamma_\rho^f(A, A) = \text{Tr } A \mathbb{J}_\rho^{-1}(A) \quad \text{and} \quad \mathbb{J}_\rho = \mathbb{R}_\rho^{1/2} f(\mathbb{L}_\rho \mathbb{R}_\rho^{-1}) \mathbb{R}_\rho^{1/2},$$

where the linear transformations  $\mathbb{L}_\rho$  and  $\mathbb{R}_\rho$  acting on matrices are the left and right multiplications, that is

$$\mathbb{L}_\rho(X) = \rho X \quad \text{and} \quad \mathbb{R}_\rho(X) = X\rho.$$

The above  $\gamma_\rho(A, A)$  is formally a quasi-entropy,  $S_{1/f}^{A\rho^{-1}}(\rho, \rho)$ , however this form is not suitable to show the monotonicity. Assume that  $\rho = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ . Then

$$\gamma_\rho^f(A, A) = \sum_{ij} \frac{1}{\lambda_i f(\lambda_j/\lambda_i)} |A_{ij}|^2. \quad (30)$$

It is clear from this formula that the Fisher information is affine in the function  $1/f$ . Therefore, **Hansen's canonical representation** of the reciprocal of a standard operator monotone function can be used [19].

**Theorem 6** *If  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  be a standard operator monotone function, then*

$$\frac{1}{f(t)} = \int_0^1 \frac{1+\lambda}{2} \left( \frac{1}{t+\lambda} + \frac{1}{1+t\lambda} \right) d\mu(\lambda),$$

where  $\mu$  is a probability measure on  $[0, 1]$ .

The theorem implies that the set  $\{1/f : f \text{ is standard operator monotone}\}$  is convex and gives the extremal points

$$g_\lambda(x) := \frac{1+\lambda}{2} \left( \frac{1}{t+\lambda} + \frac{1}{1+t\lambda} \right) \quad (0 \leq \lambda \leq 1). \quad (31)$$

One can compute directly that

$$\frac{\partial}{\partial \lambda} g_\lambda(x) = -\frac{(1-\lambda^2)(x+1)(x-1)^2}{2(x+\lambda)^2(1+x\lambda)^2}.$$

Hence  $g_\lambda$  is decreasing in the parameter  $\lambda$ . For  $\lambda = 0$  we have the largest function  $g_0(t) = (t+1)/(2t)$  and for  $\lambda = 1$  the smallest is  $g_1(t) = 2/(t+1)$ . (Note that this was also obtained in the setting of positive operator means [26], harmonic and arithmetic means.)

Via the operator  $\mathbb{J}_\rho$ , each monotone Fisher information determines a quantity

$$\varphi_\rho[A, A] := \text{Tr } A \mathbb{J}_\rho(A) \quad (32)$$

which is a quadratic cost functional. According to (29) (or Theorem 4) this possesses the monotonicity property

$$\varphi_\rho[\beta^*(A), \beta^*(A)] \leq \varphi_{\beta(\rho)}[A, A]. \quad (33)$$

Since (28) and (29) are equivalent we observe a one-to-one correspondence between monotone Fisher informations and monotone quadratic cost functions.

**Theorem 7** *If for every invertible density matrix  $\rho \in M_n(\mathbb{C})$  a positive definite sesquilinear form  $\varphi_\rho : M_n(\mathbb{C}) \times M_n(\mathbb{C}) \rightarrow \mathbb{C}$  is given such that*

- (1) *the monotonicity (33) holds for all completely positive coarse grainings  $\beta : M_n(\mathbb{C}) \rightarrow M_m(\mathbb{C})$ ,*
- (2)  *$\varphi_\rho[A, A]$  is continuous in  $\rho$  for every fixed  $A$ ,*
- (3)  *$\varphi_\rho[A, A] = \varphi_\rho[A^*, A^*]$ ,*
- (4)  *$\varphi_\rho[A, A] = \text{Tr } \rho A^2$  if  $A$  is self-adjoint and  $A\rho = \rho A$ ,*

*then there exists a unique standard operator monotone function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  such that*

$$\varphi_\rho^f[A, A] = \text{Tr } A \mathbb{J}_\rho(A)$$

*with the operator  $\mathbb{J}_\rho$  defined in Theorem 5.*

Any such cost function has the property  $\varphi_\rho[A, B] = \text{Tr } \rho A^* B$  when  $\rho$  commutes with  $A$  and  $B$ . The examples below show that it is not so generally.

**Example 6** Among the standard operator monotone functions,  $f_a(t) = (1+t)/2$  is maximal. This leads to the fact that among all monotone quantum Fisher informations there is a smallest one which corresponds to the function  $f_a(t)$ . In this case

$$F_\rho^{\min}(A) = \text{Tr } AL = \text{Tr } \rho L^2, \quad \text{where} \quad \rho L + L\rho = 2A. \quad (34)$$

For the purpose of a quantum Cramér-Rao inequality the minimal quantity seems to be the best, since the inverse gives the largest lower bound. In fact, the matrix  $L$  has been used for a long time under the name of **symmetric logarithmic derivative**, see [24] and [21]. In this example the quadratic cost function is

$$\varphi_\rho[A, B] = \frac{1}{2} \text{Tr } \rho(AB + BA) \quad (35)$$

and we have

$$\mathbb{J}_\rho(B) = \frac{1}{2}(\rho B + B\rho) \quad \text{and} \quad \mathbb{J}_\rho^{-1}(A) = \int_0^\infty e^{-t\rho/2} A e^{-t\rho/2} dt \quad (36)$$

for the operator  $\mathbb{J}$  of the previous section.

To see the second formula of (36), set  $A(t) := e^{-t\rho/2} A e^{-t\rho/2}$ . Then

$$\frac{d}{dt} A(t) = -\frac{1}{2}(\rho A(t) - A(t)\rho)$$

and

$$\int_0^\infty \frac{1}{2}(\rho A(t) + A(t)\rho) dt = [-A(t)]_0^\infty = A.$$

Hence

$$\mathbb{J}_\rho \left( \int_0^\infty A(t) dt \right) = A.$$

Let  $T = T^*$  and  $\rho_0$  be a density matrix. Then  $D(\theta) := \exp(\theta T/2)\rho_0 \exp(\theta T/2)$  satisfies the differential equation

$$\frac{\partial}{\partial \theta} D(\theta) = \mathbb{J}_{D(\theta)} T \quad (37)$$

and

$$\rho(\theta) = \frac{D(\theta)}{\text{Tr } D(\theta)} \quad (38)$$

is a kind of **exponential family**.

If  $\text{Tr } \rho_0 T = 0$  and  $\text{Tr } \rho_0 T^2 = 1$ , then

$$\frac{\partial}{\partial \theta} \text{Tr } \rho(\theta) T \Big|_{\theta=0} = 1$$

and  $T$  is a locally unbiased estimator (of the parameter  $\theta$  at  $\theta = 0$ ). Since

$$\frac{\partial}{\partial \theta} \rho(\theta) \Big|_{\theta=0} = \mathbb{J}_0(T),$$

we have equality in the Cramér-Rao inequality, see (24).  $\square$

**Example 7** The function

$$f_\beta(t) = \beta(1-\beta) \frac{(x-1)^2}{(x^\beta - 1)(x^{1-\beta} - 1)} \quad (39)$$

is operator monotone if  $0 < |\beta| < 1$ .

When  $A = i[\rho, B]$  is orthogonal to the commutator of the foot-point  $\rho$  in the tangent space, we have

$$F_\rho^\beta(A) = \frac{1}{2\beta(1-\beta)} \text{Tr} ([\rho^\beta, B][\rho^{1-\beta}, B]). \quad (40)$$

Apart from a constant factor this expression is the skew information proposed by Wigner and Yanase some time ago ([41]). In the limiting cases  $\beta \rightarrow 0$  or  $1$  we have

$$f_0(x) = \frac{x-1}{\log x}$$

and the corresponding Fisher information

$$\gamma_\rho(A, B) := \int_0^\infty \text{Tr } A(\rho + t)^{-1} B(\rho + t)^{-1} dt \quad (41)$$

is named after Kubo, Mori, Bogoliubov etc. The Kubo-Mori inner product plays a role in quantum statistical mechanics (see [12], for example). In this case

$$\mathbb{J}^{-1}(B) = \int_0^\infty (\rho + t)^{-1} B(\rho + t)^{-1} dt \quad \text{and} \quad \mathbb{J}(A) = \int_0^1 \rho^t A \rho^{1-t} dt. \quad (42)$$

Therefore the corresponding quadratic cost functional is

$$\varphi_\rho[A, B] = \int_0^1 \text{Tr } A \rho^t B \rho^{1-t} dt. \quad (43)$$

Let

$$\rho(\theta) := \frac{\exp(H + \theta T)}{\text{Tr } \exp(H + \theta T)}, \quad (44)$$

where  $\rho = e^H$ . Assume that  $\text{Tr } e^H T = 0$ . The Frechet derivative of  $e^H$  is  $\int_0^1 \text{Tr } e^{tH} T e^{(1-t)H} dt$ . Hence  $A$  is locally unbiased if

$$\int_0^1 \text{Tr } \rho^t T \rho^{1-t} A dt = 1.$$

This holds if

$$A = \frac{T}{\int_0^1 \text{Tr } \rho^t T \rho^{1-t} dt}.$$

In the Cramér-Rao inequality (22) the equality holds when  $\mathbb{J}_0(K) = \int_0^1 D^t K D^{1-t} dt$ .

Note that (44) is again an **exponential family**, the differential equation for

$$D(\theta) = \exp(H + \theta T)$$

has the form (37) with

$$\mathbb{J}_{D(\theta)}(K) = \int_0^1 D(\theta)^t K D(\theta)^{1-t} dt.$$

□

**Problem 2** It would be interesting to find more exponential families. This means solution of the differential equation

$$\frac{\partial}{\partial \theta} D(\theta) = \mathbb{J}_{D(\theta)} T, \quad D(0) = \rho_0.$$

If the self-adjoint  $T$  and the positive  $\rho$  commute, then the solution is  $D(\theta) = \exp(\theta T) \rho_0$ . A concrete example is

$$\frac{\partial}{\partial \theta} D(\theta) = D(\theta)^{1/2} T D(\theta)^{1/2}.$$

### 3.3 Manifolds of density matrices

Let  $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$  be a smooth  $m$ -dimensional manifold of invertible density matrices. When a quadratic cost function  $\varphi_0$  is fixed, the corresponding Fisher information is a Riemannian metric on the manifold. This gives a possibility for geometric interpretation of statistical statements [1, 2].

Fisher information appears not only as a Riemannian metric but as an information matrix as well. The **quantum score operators** (or logarithmic derivatives) are defined as

$$L_i(\theta) := \mathbb{J}_{\rho(\theta)}^{-1}(\partial_{\theta_i} \rho(\theta)) \quad (1 \leq i \leq m) \quad (45)$$

and

$$I_{ij}^Q(\theta) := \text{Tr } L_i(\theta) \mathbb{J}_{\rho(\theta)}(L_j(\theta)) \quad (1 \leq i, j \leq m) \quad (46)$$

is the **quantum Fisher information matrix**.

The next result is the monotonicity of Fisher information matrix.

**Theorem 8 [38]** *Let  $\beta$  be a coarse-graining sending density matrices on the Hilbert space  $\mathcal{H}_1$  into those acting on the Hilbert space  $\mathcal{H}_2$  and let  $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$  be a smooth  $m$ -dimensional manifold of invertible density matrices on  $\mathcal{H}_1$ . For the Fisher information matrix  $I^{1Q}(\theta)$  of  $\mathcal{M}$  and for Fisher information matrix  $I^{2Q}(\theta)$  of  $\beta(\mathcal{M}) := \{\beta(\rho(\theta)) : \theta \in G\}$  we have the monotonicity relation*

$$I^{2Q}(\theta) \leq I^{1Q}(\theta). \quad (47)$$

Assume that  $F_j$  are positive operators acting on a Hilbert space  $\mathcal{H}_1$  on which the family  $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$  is given. When  $\sum_{j=1}^n F_j = I$ , these operators determine a measurement. For any  $\rho(\theta)$  the formula

$$\beta(\rho(\theta)) := \text{Diag}(\text{Tr } \rho(\theta) F_1, \dots, \text{Tr } \rho(\theta) F_n)$$

gives a diagonal density matrix. Since this family is commutative, all quantum Fisher informations coincide with the classical (23) and the classical Fisher information stand on the left-hand-side of (47). The right-hand-side can be arbitrary quantum quantity but it is minimal if it based on the symmetric logarithmic derivative, see Example 6. This particular case of the Theorem is in the paper [5].

Assume that a manifold  $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$  of density matrices is given together a statistically relevant Riemannian metric  $\gamma$ . Given two points on the manifold their geodesic distance is interpreted as the statistical distinguishability of the two density matrices in some statistical procedure.

Let  $\rho_0 \in \mathcal{M}$  be a point on our statistical manifold. The geodesic ball

$$B_\varepsilon(\rho_0) := \{\rho \in \mathcal{M} : d(\rho_0, \rho) < \varepsilon\}$$

contains all density matrices which can be distinguished by an effort smaller than  $\varepsilon$  from the fixed density  $\rho_0$ . The size of the inference region  $B_\varepsilon(\rho_0)$  measures the statistical uncertainty at the density  $\rho_0$ . Following **Jeffrey's rule** the size is the volume measure determined by the statistical (or information) metric. More precisely, it is better to consider the asymptotics of the volume of  $B_\varepsilon(\rho_0)$  as  $\varepsilon \rightarrow 0$ . It is known in differential geometry that

$$Vol(B_\varepsilon(\rho_0)) = C_m \varepsilon^m - \frac{C_m}{6(m+2)} \text{Scal}(\rho_0) \varepsilon^{m+2} + o(\varepsilon^{m+2}), \quad (48)$$

where  $m$  is the dimension of our manifold,  $C_m$  is a constant (equals to the volume of the unit ball in the Euclidean  $m$ -space) and  $\text{Scal}$  means the scalar curvature, see [13, 3.98 Theorem]. In this way, the scalar curvature of a statistically relevant Riemannian metric might be interpreted as the **average statistical uncertainty** of the density matrix (in the given statistical manifold). This interpretation becomes particularly interesting for the full state space endowed by the Kubo-Mori inner product as a statistically relevant Riemannian metric.

The Kubo-Mori (or Bogoliubov) inner product is given by

$$\gamma_\rho(A, B) = \text{Tr}(\partial_A \rho)(\partial_B \log \rho), \quad (49)$$

or (41) in the affine parametrization. On the basis of numerical evidences it was **conjectured** in [35] that the scalar curvature which is a statistical uncertainty is monotone in the following sense. For any coarse graining  $\alpha$  the scalar curvature at a density  $\rho$  is smaller than at  $\alpha(\rho)$ . The average statistical uncertainty is increasing under coarse graining. Up to now this conjecture has not been proven mathematically. Another form of the conjecture is the statement that along a curve of Gibbs states

$$\frac{e^{-\beta H}}{\text{Tr } e^{-\beta H}}$$

the scalar curvature changes monotonely with the inverse temperature  $\beta \geq 0$ , that is, **the scalar curvature is monotone decreasing function of  $\beta$** . (Some partial results are in [4].)

Let  $\mathcal{M}$  be the manifold of all invertible  $n \times n$  density matrices. If we use the affine parametrization, then the tangent space  $T_\rho$  consists of the traceless self-adjoint matrices and has ab orthogonal decomposition

$$T_\rho = \{i[\rho, B] : B \in M_n^{sa}\} \oplus \{A = A^* : \text{Tr } A = 0, \quad A\rho = \rho A\}. \quad (50)$$

We denote the two subspaces by  $T_\rho^q$  and  $T_\rho^c$ , respectively. If  $A_2 \in T_\rho^c$ , then

$$F(\Delta(\rho/\rho))(A_2 \rho^{\pm 1/2}) = A_2 \rho^{\pm 1/2}$$

implies

$$\text{qCov}_\rho^f(A_1, A_2) = \text{Tr } \rho A_1^* A_2 - (\text{Tr } \rho A_1^*)(\text{Tr } \rho A_2), \quad \gamma_\rho^f(A_1, A_2) = \text{Tr } \rho^{-1} A_1^* A_2$$

independently of the function  $f$ . Moreover, if  $A_1 \in T_\rho^q$ , then

$$\gamma_\rho^f(A_1, A_2) = \text{qCov}_\rho^f(A_1, A_2) = 0.$$

Therefore, the decomposition (50) is orthogonal with respect to any Fisher information and any quadratic cost functional. Moreover, the effect of the function  $f$  and the really quantum situation are provided by the components from  $T_\rho^q$ .

### 3.4 Skew information

Let  $f$  be a standard function and  $X = X^* \in M_n$ . The quantity

$$I_\rho^f(X) := \frac{f(0)}{2} \gamma_\rho^f(\text{i}[\rho, X], \text{i}[\rho, X])$$

was called **skew information** in [19] in this general setting. The skew information is nothing else but the Fisher information restricted to  $T_\rho^q$ , but it is parametrized by the commutator.

If  $\rho = \text{Diag}(\lambda_1, \dots, \lambda_n)$  is diagonal, then

$$\gamma_\rho^f(\text{i}[\rho, X], \text{i}[\rho, X]) = \sum_{ij} \frac{(\lambda_i - \lambda_j)^2}{\lambda_j f(\lambda_i/\lambda_j)} |X_{ij}|^2.$$

This implies that the identity

$$f(0) \gamma_\rho^f(\text{i}[\rho, X], \text{i}[\rho, X]) = 2\text{Cov}_\rho(X, X) - 2\text{qCov}_\rho^{\tilde{f}}(X, X) \quad (51)$$

holds if  $\text{Tr } \rho X = 0$  and

$$\tilde{f}(x) := \frac{1}{2} \left( (x+1) - (x-1)^2 \frac{f(0)}{f(x)} \right). \quad (52)$$

The following result was obtained in [14].

**Theorem 9** *If  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is a standard function, then  $\tilde{f}$  is standard as well.*

The original proof is not easy, even matrix convexity of functions of two variables is used. Here we sketch a rather elementary proof based on the fact that  $1/f \mapsto \tilde{f}$  is linear and on the canonical decomposition in Theorem 6.

**Lemma 1** *Let  $0 \leq \lambda \leq 0$  and  $f_\lambda : \mathbb{R}^+ \rightarrow \mathbb{R}$  be a function such that*

$$\frac{1}{f_\lambda(x)} := \frac{1+\lambda}{2} \left( \frac{1}{x+\lambda} + \frac{1}{1+x\lambda} \right) = g_\lambda(x).$$

*Then the function  $\tilde{f} : \mathbb{R}^+ \rightarrow \mathbb{R}$  defined in (52) is an operator monotone standard function.*

The proof of the lemma is elementary. From the lemma and Theorem 6, Theorem 9 follows straightforwardly [40].

The skew information is the Hessian of a quasi-entropy:

**Theorem 10** *Assume that  $X = X^* \in M_n$  and  $\text{Tr } \rho X = 0$ . If  $f$  is a standard function such that  $f(0) \neq 0$ , then*

$$\frac{\partial^2}{\partial t \partial s} S_F(\rho + ti[\rho, X], \rho + si[\rho, X]) \Big|_{t=s=0} = f(0) \gamma_\rho^f(i[\rho, X], i[\rho, X])$$

for the standard function  $F = \tilde{f}$ .

The proof is based on the formula

$$\frac{d}{dt} h(\rho + ti[\rho, X]) \Big|_{t=0} = i[h(\rho), X],$$

see [40].

**Example 8** We compute the Hessian of the relative entropy of degree  $\alpha$  in an exponential parametrization:

$$\frac{\partial^2}{\partial t \partial s} S_\alpha(e^{H+tA} || e^{H+sB}) \Big|_{t=s=0} = \int_0^1 \text{Tr } e^{(1-u)H} B e^{uH} A g_\alpha(u) du,$$

where

$$g_\alpha(u) = \frac{1}{\alpha(1-\alpha)} \begin{cases} u & \text{if } 0 \leq u \leq \alpha, \\ \alpha & \text{if } \alpha \leq u \leq 1-\alpha, \\ 1-u & \text{if } 1-\alpha \leq u \leq 1 \end{cases} \quad (53)$$

for  $\alpha \leq 1/2$  and for  $\alpha \geq 1/2$   $g_\alpha = g_{1-\alpha}$ .

Since

$$\frac{\partial^2}{\partial t \partial s} S_\alpha(e^{H+tA} || e^{H+sB}) = \frac{1}{\alpha(1-\alpha)} \frac{\partial^2}{\partial t \partial s} \text{Tr } \exp \alpha(H + sB) \exp(1-\alpha)(H + tA),$$

we calculate as follows:

$$\begin{aligned} & -\frac{1}{\alpha(1-\alpha)} \text{Tr} \frac{\partial}{\partial s} \exp \alpha(H + sB) \frac{\partial}{\partial t} \exp(1-\alpha)(H + tA) \\ &= \text{Tr} \int_0^1 \int_0^1 \exp(x\alpha H) B \exp(1-x)\alpha H \exp(y(1-\alpha)H) A \exp(1-y)(1-\alpha)H dxdy \\ &= \text{Tr} \int_0^1 \int_0^1 \exp((x\alpha + (1-y)(1-\alpha))H) B \exp(((1-x)\alpha + y(1-\alpha))H) A dxdy \\ &= \text{Tr} \int_0^1 \int_0^1 \exp((x\alpha - y + y\alpha - \alpha + 1)H) B \exp(-x\alpha + y - y\alpha + \alpha)H A dxdy \\ &= \int_0^1 \int_0^1 F(-x\alpha + y - y\alpha + \alpha) dxdy \end{aligned}$$

for the functional

$$F(t) = e^{(1-t)H}Be^{tH}A.$$

We continue

$$\begin{aligned} & \int_0^1 \int_0^1 F(-x\alpha + y - y\alpha + \alpha) dx dy \\ &= \int_0^1 \int_0^1 F(-x\alpha + y(1-\alpha) + \alpha) dx dy \\ &= \int_{x=0}^1 \frac{1}{1-\alpha} \int_{z=0}^{1-\alpha} F(-x\alpha + z + \alpha) dz dx \\ &= \frac{1}{\alpha} \int_{w=-\alpha}^0 \frac{1}{1-\alpha} \int_{z=0}^{1-\alpha} F(z-w) dz dw \\ &= \int_0^1 F(u)g_\alpha(u) du, \end{aligned}$$

where  $g_\alpha$  is as above.  $\square$

$$\frac{\partial^2}{\partial t \partial s} S_\alpha(e^{H+tA+sB} || e^H) = \frac{1}{\alpha(1-\alpha)} \text{Tr} \left( \frac{\partial^2}{\partial t \partial s} \exp(1-\alpha)(H+tA+sB) \right) \exp(\alpha H),$$

We know that

$$\frac{\partial^2}{\partial t \partial s} \exp(H+tA+sB) = \Big|_{t=s=0} = \int_0^1 \int_0^s e^{(1-s)H} Be^{(s-u)H} Ae^{uH} du ds,$$

therefore

$$\frac{\partial^2}{\partial t \partial s} \exp(1-\alpha)(H+tA+sB) = (1-\alpha)^2 \int_0^1 \int_0^s e^{(1-s)(1-\alpha)H} Be^{(s-u)(1-\alpha)H} Ae^{u(1-\alpha)H} du ds,$$

therefore we obtain

$$\int_0^1 \int_0^s \text{Tr} e^{[1-(s-u)](1-\alpha)H} Be^{(s-u)(1-\alpha)H} A du ds = \int_0^1 (1-x) \text{Tr} e^{[1-x](1-\alpha)H} Be^{x(1-\alpha)H} A dx$$

If  $\alpha = 0$ , then we have the Kubo-Mori inner product.  $\square$

## 4 Von Neumann algebras

Let  $\mathcal{M}$  be a von Neumann algebra. Assume that it is in standard form, it acts on a Hilbert space  $\mathcal{H}$ ,  $\mathcal{P} \subset \mathcal{H}$  is the positive cone and  $J : \mathcal{H} \rightarrow \mathcal{H}$  is the modular conjugation. Let  $\varphi$  and  $\omega$  be normal states with representing vectors  $\Phi$  and  $\Omega$  in the positive cone. For the sake of simplicity, assume that  $\varphi$  and  $\omega$  are faithful. This means that  $\Phi$  and  $\Omega$  are

cyclic and separating vectors. The closure of the unbounded operator  $A\Phi \mapsto A^*\Omega$  has a polar decomposition  $J\Delta(\omega/\varphi)^{1/2}$  and  $\Delta(\omega/\varphi)$  is called **relative modular operator**.  $A\Phi$  is in the domain of  $\Delta(\omega/\varphi)^{1/2}$  for every  $A \in \mathcal{M}$ .

For  $A \in \mathcal{M}$  and  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ , the quasi-entropy

$$S_f^A(\omega\|\varphi) := \langle A\Phi, f(\Delta(\omega/\varphi))A\Phi \rangle \quad (54)$$

was introduced in [33], see also Chapter 7 in [31]. Of course, (5) is a particular case.

**Theorem 11** *Assume that  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is an operator monotone function with  $f(0) \geq 0$  and  $\alpha : \mathcal{M}_0 \rightarrow \mathcal{M}$  is a Schwarz mapping. Then*

$$S_f^A(\omega \circ \alpha\|\varphi \circ \alpha) \geq S_f^{\alpha(A)}(\omega\|\varphi) \quad (55)$$

holds for  $A \in \mathcal{M}_0$  and for normal states  $\omega$  and  $\varphi$  of the von Neumann algebra  $\mathcal{M}$ .

The relative entropies are jointly convex in this setting similarly to the finite dimensional case. Now we shall concentrate on the generalized variance.

## 4.1 Generalized covariance

To deal with generalized covariance, we assume that  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is a standard operator monotone (increasing) function. The natural extension of the covariance (from probability theory) is

$$\text{qCov}_\omega^f(A, B) = \langle \sqrt{f(\Delta(\omega/\omega))}A\Omega, \sqrt{f(\Delta(\omega/\omega))}B\Omega \rangle - \overline{\omega(A)}\omega(B), \quad (56)$$

where  $\Delta(\omega/\omega)$  is actually the modular operator. Although  $\Delta(\omega/\omega)$  is unbounded, the definition works. For the function  $f$ , the inequality

$$\frac{2x}{x+1} \leq f(x) \leq \frac{1+x}{2}$$

holds. Therefore  $A\Omega$  is in the domain of  $\sqrt{f(\Delta(\omega/\omega))}$ .

For a standard function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  and for a normal unital Schwarz mapping  $\beta : \mathcal{N} \rightarrow \mathcal{M}$  the inequality

$$\text{qCov}_\omega^f(\beta(X), \beta(X)) \leq \text{qCov}_{\omega \circ \beta}^f(X, X) \quad (X \in \mathcal{N}) \quad (57)$$

is a particular case of Theorem 11 and it is the monotonicity of the generalized covariance under coarse-graining. The common symmetrized covariance

$$\text{Cov}_\omega(A, B) := \frac{1}{2}\omega(A^*B + BA^*) - \overline{\omega(A)}\omega(B)$$

is recovered by the particular case  $f(t) = (1+t)/2$ .

Since

$$\text{qCov}_\omega^f(A, B) = \gamma_\omega^f(A - \omega(A)I, B - \omega(B)I),$$

it is enough to consider these sesquilinear forms on the subspace  $T_\omega := \{A \in \mathcal{M} : \omega(A) = 0\}$ .

## 4.2 The Cramér-Rao Inequality

Let  $\{\omega_\theta : \theta \in G\}$  be a smooth  $m$ -dimensional manifold in the set of normal states of the von Neumann algebra  $\mathcal{M}$  and assume that a collection  $A = (A_1, \dots, A_m)$  of self-adjoint operators is used to estimate the true value of  $\theta$ . The subspace spanned by  $A_1, A_2, \dots, A_m$  is denoted by  $V$ .

Given a standard matrix monotone function  $f$ , we have the corresponding cost function

$$\varphi_\theta[A, B] \equiv \text{qCov}_f^{\omega_\theta}(A, B)$$

for every  $\theta$  and the cost matrix of the estimator  $A$  is a positive semidefinite matrix, defined by

$$\varphi_\theta[A]_{ij} = \varphi_\theta[A_i, A_j].$$

The **bias** of the estimator is

$$\begin{aligned} b(\theta) &= (b_1(\theta), b_2(\theta), \dots, b_m(\theta)) \\ &:= (\omega_\theta(A_1 - \theta_1 I), \omega_\theta(A_2 - \theta_2 I), \dots, \omega_\theta(A_m - \theta_m I)). \end{aligned}$$

For an **unbiased estimator** we have  $b(\theta) = 0$ . From the bias vector we form a bias matrix

$$B_{ij}(\theta) := \partial_{\theta_i} b_j(\theta).$$

For a **locally unbiased estimator** at  $\theta_0$ , we have  $B(\theta_0) = 0$ .

The relation

$$\partial_{\theta_i} \omega_\theta(H) = \varphi_\theta[L_i(\theta), H] \quad (H \in V)$$

determines the **logarithmic derivatives**  $L_i(\theta)$ . The **Fisher information matrix** is

$$J_{ij}(\theta) := \varphi_\theta[L_i(\theta), L_j(\theta)].$$

**Theorem 12** *Let  $A = (A_1, \dots, A_m)$  be an estimator of  $\theta$ . Then for the above defined quantities the inequality*

$$\varphi_\theta[A] \geq (I + B(\theta)) J(\theta)^{-1} (I + B(\theta))^*$$

*holds in the sense of the order on positive semidefinite matrices.*

Concerning the proof we refer to [38].

## 4.3 Uncertainty relation

In the von Neumann algebra setting the skew information (as a sesquilinear form) can be defined as

$$I_\omega^f(X, Y) := \text{Cov}_\omega(X, Y) - \text{qCov}_\omega^{\tilde{f}}(X, Y) \tag{58}$$

if  $\omega(X) = \omega(Y) = 0$ . (Then  $I_\omega^f(X) = I_\omega^f(X, X)$ .)

**Lemma 2** Let  $\mathcal{K}$  be a Hilbert space with inner product  $\langle\langle \cdot, \cdot \rangle\rangle$  and let  $\langle \cdot, \cdot \rangle$  be a sesquilinear form on  $\mathcal{K}$  such that

$$0 \leq \langle f, f \rangle \leq \langle\langle f, f \rangle\rangle$$

for every vector  $f \in \mathcal{K}$ . Then

$$[\langle f_i, f_j \rangle]_{i,j=1}^m \leq [\langle\langle f_i, f_j \rangle\rangle]_{i,j=1}^m \quad (59)$$

holds for every  $f_1, f_2, \dots, f_m \in \mathcal{K}$ .

*Proof:* Consider the Gram matrices  $G := [\langle\langle f_i, f_j \rangle\rangle]_{i,j=1}^m$  and  $H := [\langle f_i, f_j \rangle]_{i,j=1}^m$ , which are symmetric and positive semidefinite. For every  $a_1, \dots, a_m \in \mathbb{R}$  we get

$$\sum_{i,j=1}^m (\langle\langle f_i, f_j \rangle\rangle - \langle f_i, f_j \rangle) \bar{a}_i a_j = \langle\langle \sum_{i=1}^m a_i f_i, \sum_{i=1}^m a_i f_i \rangle\rangle - \langle \sum_{i=1}^m a_i f_i, \sum_{i=1}^m a_i f_i \rangle \geq 0$$

by assumption. This says that  $G - H$  is positive semidefinite, hence it is clear that  $G \geq H$ .  $\square$

**Theorem 13** Assume that  $f, g : \mathbb{R}^+ \rightarrow \mathbb{R}$  are standard functions and  $\omega$  is a faithful normal state on a von Neumann algebra  $\mathcal{M}$ . Let  $A_1, A_2, \dots, A_m \in \mathcal{M}$  be self-adjoint operators such that  $\omega(A_1) = \omega(A_2) = \dots = \omega(A_m) = 0$ . Then the determinant inequality

$$\det([\text{qCov}_D^g(A_i, A_j)]_{i,j=1}^m) \geq \det([2g(0)I_\omega^f(A_i, A_j)]_{i,j=1}^m) \quad (60)$$

holds.

*Proof:* Let  $E(\cdot)$  be the spectral measure of  $\Delta(\omega, \omega)$ . Then for  $m = 1$  the inequality is

$$\int g(\lambda) d\mu(\lambda) \leq g(0) \left( \int \frac{1+\lambda}{2} d\mu(\lambda) - \int \tilde{f}(\lambda) d\mu(\lambda) \right),$$

where  $d\mu(\lambda) = d\langle A\Omega, E(\lambda)A\Omega \rangle$ . Since the inequality

$$f(x)g(x) \geq f(0)g(0)(x-1)^2 \quad (61)$$

holds for standard functions [16], we have

$$g(\lambda) \geq g(0) \left( \frac{1+\lambda}{2} - f(0)\tilde{f}(\lambda) \right)$$

and this implies the integral inequality.

Consider the finite dimensional subspace  $\mathcal{N}$  generated by the operators  $A_1, A_2, \dots, A_m$ . On  $\mathcal{N}$  we have the inner products

$$\langle\langle A, B \rangle\rangle := \text{Cov}_\omega^g(A, B)$$

and

$$\langle A, B \rangle := 2g(0)I_\omega^f(A, B).$$

Since  $\langle A, A \rangle \leq \langle\langle A, A \rangle\rangle$ , the determinant inequality holds, see Lemma 2.  $\square$

This theorem is interpreted as quantum uncertainty principle [3, 15, 14, 25]. In the earlier works the function  $g$  from the left-hand-side was  $(x + 1)/2$  and the proofs were more complicated. The general  $g$  appeared in [16].

## References

- [1] S. Amari, *Differential-geometrical methods in statistics*, Lecture Notes in Stat. **28** (Springer, Berlin, Heidelberg, New York, 1985)
- [2] S. Amari and H. Nagaoka, *Methods of information geometry*, Transl. Math. Monographs **191**, AMS, 2000.
- [3] A. Andai, Uncertainty principle with quantum Fisher information, J. Math. Phys. **49**, 012106 (2008).
- [4] A. Andai, Information geometry in quantum mechanics, PhD thesis, BUTE, 2004.
- [5] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**(1994), 3439–3443.
- [6] I. Csiszár, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, Magyar. Tud. Akad. Mat. Kutató Int. Közl. **8**, 85-108 (1963).
- [7] I. Csiszár, Information type measure of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar. **2**(1967), 299–318.
- [8] I. Csiszár, A class of measures of informativity of observation channels, Per. Math. Hung. **2**, 191-213 (1972).
- [9] I. Csiszár and J. Fischer, Informationsentfernungen im Raum der Wahrscheinlichkeitsverteilungen, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7**(1962), 159–180.
- [10] I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless systems*, Akadémiai Kiadó, Budapest, 1981.
- [11] W. Feller, *An introduction to probability theory and its applications*, vol. II., John Wiley & Sons, Inc., New York-London-Sydney, 1966.
- [12] E. Fick, G. Sauermann, *The quantum statistics of dynamic processes* (Springer, Berlin, Heidelberg) 1990.
- [13] S. Gallot, D. Hulin, J. Lafontaine, *Riemannian geometry*, Springer, 1993.

- [14] P. Gibilisco, D. Imparato and T. Isola, Uncertainty principle and quantum Fisher information II, *J. Math. Phys.* **48**(2007), 072109.
- [15] P. Gibilisco, D. Imparato and T. Isola, A volume inequality for quantum Fisher information and the uncertainty principle, *J. Statist.* **130**(2007), 545–559.
- [16] P. Gibilisco, F. Hiai and D. Petz, Quantum covariance, quantum Fisher information and the uncertainty principle, *IEEE Trans. Inform. Theory* **55**(2009), 439–443.
- [17] F. Hansen and G.K. Pedersen, Jensen’s inequality for operators and Löwner’s theorem, *Math. Ann.* **258**(1982), 229–241.
- [18] F. Hansen, Characterizations of symmetric monotone metrics on the state space of quantum systems, *Quantum Inf. Comput.*, **6**(2006), 597–605.
- [19] F. Hansen, Metric adjusted skew information, *Proc Natl Acad Sci USA*. **105**(2008), 9909–9916.
- [20] M. Hayashi, *Quantum information. An introduction*, Springer-Verlag, Berlin, 2006.
- [21] C. W. Helstrom, *Quantum detection and estimation theory*, Academic Press, New York, 1976.
- [22] F. Hiai, M. Ohya and M. Tsukada, Sufficiency, KMS condition and relative entropy in von Neumann algebras, *Pacific J. Math.* **96**(1981), 99–109.
- [23] F. Hiai and D. Petz, Riemannian geometry on positive definite matrices related to means, *Lin. Alg. Appl.* **430**(2009), 3105–3130.
- [24] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, North-Holland, Amsterdam, 1982.
- [25] H. Kosaki, Matrix trace inequality related to uncertainty principle, *Internat. J. Math.* **16**(2005), 629–645.
- [26] F. Kubo and T. Ando, Means of positive linear operators, *Math. Ann.* **246**(1980), 205–224.
- [27] S. Kullback and R.A. Leibler, On information and sufficiency. *Ann. Math. Statistics*, **22**(1951), 79–86.
- [28] S. Kullback, *Information theory and statistics*, John Wiley and Sons, New York; Chapman and Hall, Ltd., London, 1959.
- [29] E. H. Lieb, Convex trace functions and the Wigner-Yanase-Dyson conjecture, *Advances in Math.* **11**(1973), 267–288.
- [30] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, *IEEE Trans. Inform. Theory* **52**(2006), 4394–4412.

- [31] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993. Second edition 2004.
- [32] F. Österreicher and I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, Ann. Inst. Statist. Math. **55**(2003), 639–653.
- [33] D. Petz, Quasi-entropies for states of a von Neumann algebra, Publ. RIMS. Kyoto Univ. **21**(1985), 781–800.
- [34] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys., **23**(1986), 57–65.
- [35] D. Petz, Geometry of Canonical Correlation on the State Space of a Quantum System, J. Math. Phys. **35**(1994), 780–795.
- [36] D. Petz, Monotone metrics on matrix spaces, Linear Algebra Appl. **244**(1996), 81–96.
- [37] D. Petz and Cs. Sudár, Geometries of quantum states, J. Math. Phys. **37**(1996), 2662–2673.
- [38] D. Petz, Covariance and Fisher information in quantum mechanics. J. Phys. A: Math. Gen. **35**(2003), 79–91.
- [39] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer, Berlin, Heidelberg, 2008.
- [40] D. Petz and V.E.S. Szabó, From quasi-entropy to skew information, preprint, 2007, to be published in Int. J. Math.
- [41] E.P. Wigner, M.M. Yanase, Information content of distributions, Proc. Nat. Acad. Sci. USA **49**(1963), 910–918.